Registered Report: Exploratory Analysis of Ownership Diversity and Innovation in the Annual Business Survey

by

Timothy R. Wojan
National Center for Science and Engineering Statistics - National Science Foundation

CES 23-11 **March 2023**

Abstract

A lack of transparency in specification testing is a major contributor to the replicability crisis that has eroded the credibility of findings for informing policy. How diversity is associated with outcomes of interest is particularly susceptible to the production of nonreplicable findings given the very large number of alternative measures applied to several policy relevant attributes such as race, ethnicity, gender, or foreign-born status. The very large number of alternative measures substantially increases the probability of false discovery where nominally significant parameter estimates—selected through numerous though unreported specification tests—may not be representative of true associations in the population. The purpose of this registered report is to: 1) select a single measure of ownership diversity that satisfies explicit, requisite axioms; 2) split the Annual Business Survey (ABS) into an exploratory sample (35%) used in this analysis and a confirmatory sample (65%) that will be accessed only after the publication of this report; 3) regress self-reported new-to-market innovation on the diversity measure along with industry and firm-size controls; 4) pass through those variables meeting precision and magnitude criteria for hypothesis testing using the confirmatory sample; and 5) document the full set of hypotheses to be tested in the final analysis along with a discussion of the false discovery and family-wise error rate corrections to be applied. The discussion concludes with the added value of implementing split sample designs within the Federal Statistical Research Data Center system where access to data is strictly controlled.

Keyword*:* Split-sample, false discovery, self-reported innovation, women and minority owned business, hypothesis testing

JEL Classification: O3, J15, J16, C1

---

[*] **Disclaimer**: Any views expressed are those of the author and not those of the U.S. Census Bureau, the National Center for Science and Engineering Statistics, or the National Science Foundation. The Census Bureau's Disclosure Review Board has reviewed this information product for unauthorized disclosure of confidential information and have approved the disclosure avoidance practices applied to this release. This research was performed at a Federal Statistical Research Data Center under FSRDC Project Number 2681 and at the National Center for Science and Engineering Statistics Survey Sponsor Data Center (Approval ID: CBDRB-FY23-0188, CBDRB-FY20-008). The author has no competing interests to declare that are relevant to the content of this article.

Acknowledgement: This work was supported in part by the National Science Foundation (NSF), National Center for Science and Engineering Statistics (NCSES), and by the Oak Ridge Institute for Science and Education (ORISE) for the Department of Energy (DOE). ORISE is managed by Oak Ridge Associated Universities (ORAU) under DOE contract number DE-SC0014664. All opinions expressed in this paper are the author's and do not necessarily reflect the policies and views of NSF, DOE, ORAU, or ORISE.

Corresponding author:
Timothy R. Wojan, ORISE Established Scientist Fellow
National Center for Science and Engineering Statistics within the National Science Foundation
Alexandria, VA
Email: twojan@nsf.gov

Introduction

A common practice in applied research is to use the same dataset for both specification testing to select a model, and for hypothesis testing of whether the phenomenon of interest is likely to be present in the population. The consequence of this practice is to invalidate the hypothesis test statistics from statistical software packages that assume that the test is *de novo* with no prior testing or selection. The same problem arises in comparing the statistical significance of alternative proxies to represent a phenomenon. A nominal p-value of 0.05 for a single variable increases to a true p-value of 0.14 if three variables are compared and to 0.46 if 12 variables are compared.[1] False discovery rate or family-wise error rate corrections can be used to transform nominal p-values produced by statistical software into true p-values, but these are rarely applied (Ferraro and Shukla 2020). Null hypothesis statistical testing that dominates in applied research provides perverse incentives to report nominal p-values as true p-values, and to regard specification testing as having minimal impact on the statistical validity of published findings.

Two methods to address this lack of transparency in applied research are 1) publishing pre-analysis plans that "tie the hands" of researchers by prohibiting specification testing to ensure that hypothesis tests are in fact *de novo* (Casey, et al. 2012)*;* and 2) split sample designs that limit specification testing to an exploratory subsample with final models tested *de novo* using a confirmatory subsample (Anderson and Magruder 2017; Heller, et al. 2009). The latter method is used in this analysis given the very large size of the Annual Business Survey (ABS) that guarantees statistically powerful tests even after splitting the sample. The advantage of split sample design is the ability to explore for possible statistical associations and then to test the robustness of identified associations. This ability is invaluable for examining the association between ownership diversity and innovation in the ABS given the very large number of ways that diversity may be defined and measured. Diversity can be defined on a single attribute such as race or gender or as a composite of multiple attributes. ABS contains 7 attributes of owners that can be used to define diversity including race, ethnicity, foreign-born status, gender, age, level of education, and educational specialization. Diversity among any of these attributes might increase the probability of self-reported innovation given the different perspectives and life experiences available to an ownership team. Alternatively, homophily may increase the probability of self-reported innovation if the flow of novel information is facilitated among "birds of a feather" (Luo and Deng 2009). How diversity and homophily affect each attribute or the interaction of attributes over 120 possible combinations presents both opportunities for discovery as well as opportunities for false discovery.

The opportunities for false discovery multiply substantially with the alternatives available for measuring diversity. One strategy for arriving at a single, "best," diversity measure to reduce the dimensionality of this problem is to select a diversity measure that satisfies requisite axioms defined by the researcher. This removes the temptation to add index searches as part of specification testing, selecting diversity measures that comport with one's priors inductively. While the axiomatic selection does not definitively resolve problem of the most appropriate diversity measure to use, it at least makes the defense of the measure used transparent.

The discussion begins with the axiomatic selection of a single diversity measure to be used in all analyses. The justification and procedures for splitting the sample into exploratory and confirmatory

[1] If the null is true in the population, then the probability of falsely rejecting the null *Pr*(at least 1 rejection) = 1 − *Pr*(all accepted) = 1 − $(1 - \alpha)^k$ where k is the number of variables tested.

subsamples is provided before the specification of the logistic regression equation examining the association between self-reported innovation and ownership diversity is discussed. Estimates are provided and discussed with respect to the single attribute diversity measures and the composite diversity measures. The documentation of exploratory hypotheses passed through for later confirmatory tests is accompanied by a discussion of the required Bonferroni and false discovery rate corrections to be applied. The paper concludes with a discussion of the advantages of conducting split sample design studies on secondary data within the Federal Statistical Research Data Center system.

Selecting a Single Diversity Measure Axiomatically

The wide range of possible diversity measures makes any analysis of the association of diversity with innovation susceptible to false discovery. A researcher could compute many diversity measures, test the association of each measure with some innovation indicator, and then select the diversity measure producing the result that most strongly reinforced the researcher's priors, treating the nominal p-value as a true p-value despite repeated specification testing. The approach adopted here to guard against false discovery and produce findings with a high probability of being replicable is to arrive at a diversity measure satisfying axioms stipulated at the beginning of the research before any associations between diversity and innovation are estimated (Wojan 2022). The axiomatic approach to poverty measurement first proposed by Sen (1976) is adapted to diversity measurement as applied to the case of diversity among small ownership teams. Data on owner characteristics in the Annual Business Survey is only available for up to four owners along the dimensions of gender, race, ethnicity, foreign-born status, age, level of education, and area of educational specialization. Data on the share of the business claimed by each owner is also available. A measure that incorporates information on homophily or the degree of fractionalization, the influence of concentrated or distributed owner shares, and the number of owners likely to reflect the potential diversity of ideas applied to innovation challenges should satisfy the following axioms:

HOMOPHILY AXIOM: Given other things, all owners belonging to the same group must result in the lowest diversity measure value.

FRACTIONALIZATION AXIOM: Given other things, an increase in the number of groups must increase the diversity measure value.

TEAM SIZE AXIOM: Given other things, larger ownership teams not demonstrating homophily must increase the diversity measure value relative to smaller ownership teams.

CONCENTRATION OF OWNERSHIP AXIOM: Given other things, ownership concentrated in one member of the team must reduce the diversity measure value relative to ownership that is more equally distributed among team members.

The homophily axiom defines the standard for comparing different degrees of diversity among ownership teams. As such, the value of the diversity measure must always be lowest for those ownership teams that lack diversity in the dimension of interest. This does raise the issue of whether firms with single owners, for which diversity along any dimension or across many dimensions is impossible, should be excluded from the analysis. Since the main research interest is comparing the association of homophilic and heterophilic collaboration with innovation, single owner firms are excluded from the analysis as their inclusion would confound any effect from the lack of diversity on innovation with the effect of no owner collaboration on innovation.

The fractionalization axiom requires that the diversity measure must increase with the number of unique groups in any ownership team dimension. The number of unique groups cannot exceed four for any given dimension which would be the case if each owner in a four-owner firm belonged to a different group. Ownership team size will thus pose a limit on the number of unique groups in any dimension. However, a maximum of two groups characterizes dimensions defined as binary in the 2018 ABS such as gender or foreign-born status.

The team size axiom is arguably the most controversial as it conflicts with the common assumption that diversity is a function of relative composition defined by population shares, not dependent on population size. And this is where the construct of diversity as a population characteristic breaks down for understanding diversity in ownership teams. The probability of interaction within small ownership teams is 1 whereas the probability of members of one group interacting with members of another in a population is always less than 1 but governed by the relative size of shares. The issue is whether guaranteed interaction with more members of another group should result in a higher measure of diversity than guaranteed interaction with fewer members. To make this concrete, consider an ownership team of two that has members from different groups versus an ownership team of 4 that are members from 2 different groups. From a population share perspective the level of diversity is identical. However, owners in the 4-owner firm will each be exposed to 2 viewpoints from another group versus the single different viewpoint available in the 2-owner firm. If the ownership team is conceived as a network, then the number of diverse nodes increases with the size of the network for all heterophilic teams.

The other salient difference between a population and ownership team is the separation of the ownership interest from the characteristics of any individual owner. In contrast to "one person, one vote," the relative power or influence of any owner may be dominant or nominal. The concentration of ownership axiom recognizes that the consideration of diverse viewpoints is likely to be greater in a firm characterized by equal shares of ownership relative to the firm where one owner controls a dominant interest. Ownership shares can be treated as population shares found in more traditional diversity measures.

The commonly used ethno-linguistic fractionalization (ELF) index is computed as

$$ELF = 1 - \sum_{i=1}^{n} p_i^2 \qquad \textit{Equation 1}$$

where *p* represents population or ownership shares of *n* different groups. The index satisfies three of the four axioms but is invariant to population size so does not satisfy the team size axiom.

A minor modification of the ELF index that satisfies all four axioms is produced by replacing the square term of the population shares with the number of unique groups (*n*) as the share exponent, summed over the total number of owners (*o*)

$$OF = 1 - \sum_{i=1}^{o} p_i^n \qquad Equation\ 2$$

where *p* represents the ownership share of the i^{th} owner.

Tables showing the range of values for the ownership fractionalization (OF) index for all possible group and owner combinations, for both highly concentrated and equally distributed business ownership, and the values that would result from applying the ELF index are provided below.

Table 1 OF Index

	Unique Groups						
	Homophily	2	2	3	3	4	4
Owners		Concentrated	Equally Distributed	Concentrated	Equally Distributed	Concentrated	Equally Distributed
2	0	0.0198	0.5	0.0394	0.6667	0.0588	0.75
3	0			0.0588	0.8889	0.0873	0.9375
4	0					0.1147	0.9844

Table 2 ELF Index

	Unique Groups						
	Homophily	2	2	3	3	4	4
Owners		Concentrated	Equally Distributed	Concentrated	Equally Distributed	Concentrated	Equally Distributed
2	0	0.0198	0.5	0.0394	0.6667	0.0588	0.75
3	0			0.0394	0.6667	0.0589	0.75
4	0					0.0588	0.75

Invariance to the size of the ownership team using the ELF index is clear, as is the smaller range of the ELF index relative to the OF index. When applied to ABS data the OF index will also result in larger variance across observations which is a desirable characteristic for an independent variable tasked with explaining variance in a dependent variable.

Finding an index that satisfies the required axioms for analyzing the potential impact of ownership diversity on innovation resolves the most serious problem of false discovery resulting from index searches as part of specification searches. However, the number of dimensions that are of interest with respect to diversity and innovation is significantly greater than 1, with the implication that the multiple comparison problem has not been fully resolved. Consideration of composite diversity measures that combine diversity from multiple dimensions greatly increases the possible number of comparisons. Diversity across multiple dimensions is how diversity is experienced in the real world so exploratory tests of composite diversity indices should be part of the first stage of the analysis. The number of unique combinations of the 7 different dimensions of diversity is 120. An unweighted composite ownership fractionalization (COF) index can be expressed as

$$COF = \frac{\left(D - \sum_{i=1}^{D} \sum_{j=1}^{o} p_j^n\right)}{D} \qquad Equation\ 3$$

where *D* represents the number of dimensions included in the index. Normalizing by D ensures that the magnitude of effects measured via odds ratios for both the COF and OF indices will be comparable.

Split Sample Procedures

Anderson and Magruder (2017) examine various trade-offs in the design of split-sample studies regarding the statistical power available in the confirmatory sample, the threshold for passing exploratory hypotheses on for confirmation, and the use of one-sided tests. The statistical power implications of split sample design are critical in studies using primary data where sample sizes tend to be small, and these implications are investigated using Monte Carlo methods. Their simulations suggest that a 35%/65% split between exploratory and confirmatory samples, respectively, is ideal. This guidance is followed in this exploratory analysis. While the large sample size of the ABS reducing concerns about statistical power, it does raise concerns regarding the threshold for passing exploratory hypotheses on for confirmation. Large sample sizes can produce results that are statistically different from zero but that are still very close to zero. That is, the statistical test may detect a distinction without a substantive difference. To ensure that exploratory hypotheses that are passed through for confirmation are economically significant, a pass-through threshold for magnitude will also be applied. In addition to being statistically significant at the conventional 0.05 level, parameter estimates will also be required to meet a minimum effect size corresponding to a Cohen's d of ≥ 0.2 or "small effect," which is equivalent to an odds ratio of ≥ 1.44, or < 0.6945 (Borenstein, et al. 2021). All parameter estimates from the exploratory analysis are provided below if readers want to investigate the magnitude of statistically significant parameters estimates that are not passed through for confirmation. Finally, all exploratory tests will use two-sided tests given uncertainty regarding the direction of association of diversity or homophily on innovation.

Specification of the Innovation and Ownership Diversity Regression Equation

The central interest of investigating the association between ownership diversity and innovation requires specifying a regression equation that controls for alternative sources of variation that might also be correlated with diversity. For example, larger firms are more likely to report innovation, but larger firms are also more likely to have larger ownership teams that opens the possibility of greater diversity. For this reason, firm size categories controlling for this source of variation should be included in the regression equation. Similarly, some industries have a higher percentage of family-owned businesses—such as Construction, and Accommodation and Food Services—that are less likely to report innovation and will also tend to be less diverse across several dimensions. Including two-digit NAICS controls in the regression equation will provide better estimates of the independent effect of diversity on innovation. Other controls could be included but industry and firm size are noncontroversial and provide the parsimonious specification:

$$\ln\left[\frac{y}{1-y}\right] = \beta_0 + \beta_1 x_1 + FE_{firmsize} + FE_{industry} + \varepsilon \qquad \textit{Equation 4}$$

where y = self-reported new-to-market innovation;
x_1 = OF or COF index;
$FE_{firmsize}$ = firm size fixed effects array;
$FE_{indsutry}$ = industry fixed effects array;
ε = error term.

There are several dependent variables that could be used but research by Tian, et al. (2022) demonstrate that innovation that is "new to the market" is more likely to represent more far-ranging, novel innovation and less likely to include incremental innovation. The response to the new-to-market innovation question is thus the most appropriate for assessing the extent to which diversity may either spawn or inhibit more novel ideas.

Estimates of the Association of Single Attribute Diversity with Innovation

The report of exploratory findings begins with the single attribute diversity measure (OF) as these results are simplest to interpret and provide clues on how the composite diversity measures may affect innovation. The comparison of estimates for racial or gender diversity with diversity in educational specialization provides a clear indication of whether innovation as a combination of ideas from different academic domains extends to the combination of ideas from different lived experiences.

Descriptive statistics in Table 3 of single attribute diversity measures provides information on the distribution of various types of ownership diversity in the business population. Ownership teams are most likely to demonstrate gender, educational level, and age diversity. The relatively high rate of diversity among gender and age may be reflective of family-owned businesses or jointly owned by spouses/partners which make up 32.4% of the exploratory sample. In contrast, diversity across race, ethnicity, foreign-born status, or educational specialization is relatively rare, with more than 90% of ownership teams being homophilic along each of these dimensions.

Table 3 Descriptive Statistics of Single Attribute Ownership Diversity, Innovation, and Ownership Status

Attribute	Mean	Median	Percent Homophilic
Age	0.2544	0	52.87
Education	0.2894	0.3674	44.61
Gender	0.326	0.4998	34.2
Ethnicity	0.0212	0	95.8
Educational Specialization	0.0472	0	91.31
Race	0.0228	0	95.64
Foreign-born Status	0.0465	0	90.86
New-to-Market Innovation	0.0895	0	N/A
Family-owned Businesses	0.3037	0	N/A
Jointly owned by Spouses/Partners	0.2284	0	N/A
Neither Family nor Jointly Owned	0.6761	1	N/A

Source: 2018 Annual Business Survey, 35% Exploratory Sample

The association between single attribute ownership diversity with new-to-market-innovation, using the ownership fractionalization measure and controlling for industry and firm size, is presented in Table 4. Gender is the only attribute that is not statistically significant in its association with self-reporting new-to-market innovation, and thus is not passed through for confirmatory testing. The coefficient estimate for Age is statistically significant but fails to pass the magnitude criterion with an odds ratio of only 1.13 at the point estimate. Of the 5 single attribute diversity measures passed through, educational specialization has the largest magnitude, reinforcing the prior that innovation may be more common among teams coming from different disciplines. The coefficient estimates for Race and Foreign-born Status are smaller in magnitude but provide preliminary evidence suggesting that diversity in lived experience may also increase the likelihood of innovation within a business. Diversity in the level of Education or the Ethnicity of owners are also passed through for confirmatory testing, meeting both statistical significance and magnitude criteria.

Table 4 Association Between Single Attribute Diversity and New-to-Market Innovation

Attribute	Estimate	Standard Error	Odds Ratio	Passed Through?
Age	0.1219	0.0141	1.13	No
Education	0.4189	0.0144	1.52	Yes
Gender	-0.0024	0.0163	0.998	No
Ethnicity	0.4542	0.0368	1.575	Yes
Educational Specialization	0.7175	0.0145	2.049	Yes
Race	0.5619	0.0333	1.754	Yes
Foreign-born Status	0.5603	0.0246	1.751	Yes

Source: 2018 Annual Business Survey, 35% Exploratory Sample
Note: Coefficient estimates for firm size class and industry controls are not shown.

Estimates of the Association of Composite Diversity with Innovation

Estimates from the single attribute diversity measure regression equation essentially restrict any influence from composite diversity (e.g., owners of different races and different gender) to be zero. The validity of this assumption can be investigated empirically by estimating the full complement of composite diversity measures across all 7 attributes. The composite measures are a better representation of how diversity is experienced in the real world as interaction is between holistic actors rather than between their isolated attributes. However, if some diversity among some attributes facilitates innovation while diversity among other attributes is neutral or impedes innovation then the composite measures may result in measures that are smaller in magnitude or estimated with less precision. Alternatively, if some forms of diversity over multiple attributes are synergistic then those estimates would tend to be larger.

Table 5 shows the 9 combinations passed through for confirmatory testing out of a possible 21 unique dual combinations across 7 attributes. The most surprising result is that none of the combinations passed through include Educational Specialization, despite this attribute having the largest magnitude in the single attribute estimation. One possible explanation for this is that Educational Specialization is not defined for owners having less than a 4-year college degree. For teams with both college graduates and nongraduates this is not a problem as nongraduates can be classified as a unique specialization group.

Similarly, ownership teams with no college graduates can be classified as homophilic on this dimension. A robustness check excluding homophilic nongraduate teams did not qualitatively change the results. Investigating the role of Educational Specialization diversity within knowledge intensive or R&D performing businesses is a topic for future research but is not pursued in this exploratory analysis.

Table 5 Dual Attribute Diversity and New-to-Market Innovation Passed Through for Confirmatory Testing

Attribute	Estimate	Standard Error	Odds Ratio
Race + Foreign-born	0.899	0.035	2.433
Ethnicity + Race	0.9432	0.0372	2.568
Ethnicity + Foreign-born	0.9012	0.0376	2.463
Education + Foreign-born	0.8256	0.0236	2.283
Education + Race	0.8578	0.0261	2.358
Education + Ethnicity	0.8201	0.0263	2.271
Education + Gender	0.3928	0.0198	1.481
Age + Foreign-born	0.4043	0.0233	1.498
Age + Education	0.3997	0.0173	1.491

Source: 2018 Annual Business Survey, 35% Exploratory Sample
Note: Coefficient estimates for firm size class and industry controls are not shown.

The most common attribute included in the combinations passed through for confirmation is diversity in the level of Education, appearing in 5. Foreign-born Status appears in 4 of the combinations passed through, with Ethnicity and Race included in 3, Age included in 2, and Gender included in 1. Combinations including Race, Ethnicity, and Foreign-born Status also tend to have the largest magnitude.

Table 6 shows the 17 combinations passed through for confirmatory testing out of 35 unique combinations of 3 elements across 7 attributes. Again, we see that Educational Specialization is not included in any of the combinations passed through. Education and Foreign-born Status appear in 10 of the combinations, with Age appearing in 8, and seven combinations include Race, Ethnicity, and Gender. The magnitudes for some of the combinations with Ethnicity, Race, or Foreign-born Status are considerably larger than the other 3-attribute combinations. For example, the Ethnicity + Education + Race coefficient estimate can be interpreted as follows: a 4-person ownership team characterized by equal ownership and 4 races, 4 educational levels, and both Hispanic and non-Hispanic members would be roughly 3 times as likely to report innovation as a 4-person team that was homophilic across these attributes.[2]

[2] The odds ratio is based on a 1-unit change in the independent variable and the composite ownership fractionalization measure including one binary attribute (Ethnicity) has a maximum value of 0.823.

Table 6 Three Attribute Diversity and New-to-Market Innovation Passed Through for Confirmatory Testing

Attribute	Estimate	Standard Error	Odds Ratio
Ethnicity + Race + Foreign-born	1.179	0.0444	3.253
Gender + Race + Foreign-born	0.5768	0.0361	1.78
Gender + Ethnicity + Foreign-born	0.5313	0.0369	1.701
Education + Race + Foreign-born	0.4183	0.0406	1.519
Education + Ethnicity + Foreign-born	1.168	0.0317	3.216
Education + Ethnicity + Race	1.234	0.0359	3.435
Education + Gender + Foreign-born	0.6871	0.0267	1.988
Education + Gender + Race	0.6613	0.0283	1.937
Education + Gender + Ethnicity	0.6217	0.0283	1.862
Age + Race + Foreign-born	0.6467	0.0313	1.909
Age + Ethnicity + Foreign-born	0.6308	0.0322	1.879
Age + Ethnicity + Race	0.589	0.0354	1.802
Age + Gender + Foreign-born	0.3802	0.0279	1.463
Age + Education + Foreign-born	0.6523	0.0234	1.92
Age + Education + Race	0.6443	0.0248	1.905
Age + Education + Ethnicity	0.6229	0.25	1.864
Age + Education + Gender	0.4163	0.0217	1.516

Source: 2018 Annual Business Survey, 35% Exploratory Sample
Note: Coefficient estimates for firm size class and industry controls are not shown.

Table 7 shows the 15 combinations passed through for confirmation out of 35 unique combinations of 4 elements across 7 attributes. Age, Education, Gender, Ethnicity, Race, and Foreign-born Status appear in 10 of the combinations with Educational Specialization appearing in none. The coefficient estimate for the combination made up of Education + Ethnicity + Race + Foreign-born Status is the largest in magnitude of any of the 127 possible estimates. Given a maximum possible range of 0.7422 for this combination with 2 binary attributes, an ownership team as diverse as possible along these dimensions would be more than 3 times as likely to report new-to-market innovation compared to a homophilic ownership team.

Table 7 Four Attribute Diversity and New-to-Market Innovation Passed Through for Confirmatory Testing

Attribute	Estimate	Standard Error	Odds Ratio
Gender + Ethnicity + Race + Foreign-born	0.801	0.044	2.228
Education + Ethnicity + Race + Foreign-born	1.466	0.039	4.333
Education + Gender + Race + Foreign-born	0.9438	0.0334	2.57
Education + Gender + Ethnicity + Foreign-born	0.917	0.0337	2.502
Education + Gender + Ethnicity + Race	0.9043	0.0359	2.47
Age + Ethnicity + Race + Foreign-born	0.8716	0.0387	2.391
Age + Gender + Race + Foreign-born	0.5811	0.0345	1.788
Age + Gender + Ethnicity + Foreign-born	0.5522	0.0351	1.737
Age + Gender + Ethnicity + Race	0.4876	0.0375	1.628
Age + Education + Race + Foreign-born	0.883	0.0295	2.418
Age + Education + Ethnicity + Foreign-born	0.8721	0.0299	2.392
Age + Education + Ethnicity + Race	0.8762	0.0317	2.402
Age + Education + Gender + Foreign-born	0.6339	0.0268	1.885
Age + Education + Gender + Race	0.6115	0.028	1.843
Age + Education + Gender + Ethnicity	0.5859	0.0281	1.797

Source: 2018 Annual Business Survey, 35% Exploratory Sample
Note: Coefficient estimates for firm size class and industry controls are not shown.

Table 8 shows the 6 combinations passed through for confirmation out of 21 unique combinations of 5 elements across 7 attributes. Age, Education, Gender, Ethnicity, Race, and Foreign-born Status each appear in 5 of the combinations and Educational Specialization appears in none. One notable characteristics of Table 8 is that all of the odds ratios are greater than 2 that was not the case for any earlier table. In addition, the odds ratios are larger than any of the odds ratios included in the single diversity attribute table (Table 4). The implication is that composite diversity measures may do a better job of estimating the effects of diversity on innovation than single attribute diversity, which assumes that the impact from all other forms of diversity is zero.

Table 8 Five Attribute Diversity and New-to-Market Innovation Passed Through for Confirmatory Testing

Attribute	Estimate	Standard Error	Odds Ratio
Education + Gender + Ethnicity + Race + Foreign-born	1.168	0.0396	3.215
Age + Gender + Ethnicity + Race + Foreign-born	0.7609	0.0409	2.14
Age + Education + Ethnicity + Race + Foreign-born	1.098	0.0353	2.997
Age + Education + Gender + Race + Foreign-born	0.827	0.0321	2.287
Age + Education + Gender + Ethnicity + Foreign-born	0.8081	0.0323	2.244
Age + Education + Gender + Ethnicity + Race	0.792	0.0339	2.208

Source: 2018 Annual Business Survey, 35% Exploratory Sample
Note: Coefficient estimates for firm size class and industry controls are not shown.

Table 9 shows the single combination passed through for confirmation out of 7 unique combinations of 6 elements across 7 attributes. The combination passed through is the one of seven that does not include Educational Specialization. The magnitude of the coefficient estimate reinforces the preliminary finding that composite diversity measures may better capture the phenomenon of interest. The exploratory nature of this analysis cautions against any definitive statements. The confirmatory testing that will include family-wise error rate and false discovery rate corrections for multiple comparisons in the next stage of the analysis is outlined below.

Table 9 Six Attribute Diversity and New-to-Market Innovation Passed Through for Confirmatory Testing

Attribute	Estimate	Standard Error	Odds Ratio
Age + Education + Gender + Ethnicity + Race + Foreign-born	1.002	0.0372	2.724or

Source: 2018 Annual Business Survey, 35% Exploratory Sample
Note: Coefficient estimates for firm size class and industry controls are not shown.

Hypothesis Tests to Be Conducted Using Confirmatory Sample

Tables 4 – 9 document the specific variables to be tested using the confirmatory sample once this Registered Report is published. However, comparison of several parameter estimates is still vulnerable to false discovery and this probability increases with the number of tests. There are two ways of correcting the nominal p-values provided by statistical software to represent true p-values. The more conservative method is the family wise error rate correction which essentially sets the error rate for a single false discovery over the entire collection of parameter estimates being compared. It is derived by dividing the desired p-value by the number of estimated parameters being compared (*m)*. The nominal p-value for each individual estimate to achieve $\alpha < 0.05$ for the collection of estimates would thus be α/m. Since 57 parameters were passed through, the nominal p-values required for significance using the Bonferroni correction would be $p = 0.000877$. If the number of parameter estimates to be compared is large, the Bonferroni correction can make rejection of the null highly improbable, increasing the likelihood of Type II (false negative) errors. For example, without the elimination of more than half of the composite measures from the list of variables passed on for confirmatory testing, the Bonferroni correction would require $p = 0.00039$ (0.05/127). In this way, split sample designs can increase the statistical power of hypothesis tests by potentially reducing the number of comparisons resulting in less stringent corrections and thus fewer false negatives. That is, statistical power is increasing in the required p threshold.

The false discovery rate correction is less conservative, requiring that the average false discovery rate of the entire collection of parameter estimates is equal to α. This correction requires ranking the m parameter estimates from most precise to least precise and then setting the nominal p-value to satisfy the significance threshold for the i^{th} variable as

$$p(i) \leq \alpha \times i/m \qquad \textit{Equation 5}$$

The only parameter estimate subject to the more stringent family-wise error rate correction is the one which is most precise. For the least precise estimate to be considered statistically significant, it would only have to have p equal to the nominal α. Both the false discovery and family wise error rate corrections will be applied to the confirmatory hypothesis tests to provide researchers with the strength of evidence that the associations between ownership diversity and innovation are representative of the population.

Split Sample Proposals Using Restricted Data

The development of registries for pre-analysis plans in applied research using primary data has far outpaced the development of institutional mechanisms for verifying the transparency of split sample protocols using secondary data (Miguel, et al. 2014; Banerjee, et al 2020). In those cases where secondary data is publicly available, no such mechanisms are likely to be possible and any research transparency claimed would be dependent on the affirmation by researchers. This is the process currently used by the Center for Open Science in their Secondary Data Preregistration Wiki included in their Open Science Framework (https://osf.io/x4gzt/wiki/Preregistration%20Template/). The template requires submitting information on prior use of the data, at what point in the research process preregistration is sought, and any anticipated exploratory analysis. The template does not explicitly ask about protocols used in the split sample design, but this information could be supplied and would be a part of the publicly available registration. Supplying this information prior to analysis would be a huge improvement in the transparency surrounding applied research using secondary data. However, there is little incentive for researchers to unilaterally register analyses that would prohibit common questionable research practices, such as selecting final control variables after looking at preliminary results, or HARKing (hypothesizing after results are known), that may increase the probability of a manuscript being accepted for publication (Ferraro and Shukla 2022). And from the journal editors' perspective, there is little reason to accord special status to manuscripts purporting greater transparency using registered reports that may be produced as *post hoc* cover for questionable research practices.

This is where the highly monitored and regulated access to data in the Federal Statistical Research Data Centers may provide the opportunity to add independent, institutional guarantees that the hypothesis tests using a confirmatory sample are in fact *de novo*. Because access to confidential data is only granted after approval of a research proposal through the Standard Application Process, it would be possible to limit initial access to an exploratory sample. After a registered report for the exploratory analysis is published, access to the confirmatory sample could be provided. The central problem of verifying that hypothesis tests in the final analysis conform with a pre-analysis plan prohibiting additional specification tests is resolved. The largest cost of this approach is additional data storage required for retaining separate exploratory and confirmatory samples in addition to the source data files. Other complications may arise if a research project requires merging respondent records from several different datasets.

However, these costs and complications appear minor relative to the possible undermining of research output lacking full transparency, regardless of the validity of the criticism. Even more so if the topic being studied is contentious as is the case of the present exploratory analysis of diversity, extending to climate change, tax policy, foreign trade, labor saving technologies such as robotics or artificial intelligence, among others. In an era when widespread encouragement to “do your own research” is code for “don’t trust science,” the premium for research transparency has never been higher.

References

Anderson, M.L. and Magruder, J. 2017. *Split-sample strategies for avoiding false discoveries*. NBER Working Paper No. 23544. https://www.nber.org/papers/w23544

Banerjee, A., Duflo, E., Finkelstein, A., Katz, L.F., Olken, B.A. and Sautmann, A., 2020. *In praise of moderation: Suggestions for the scope and use of pre-analysis plans for rcts in economics* (No. w26993). National Bureau of Economic Research.

Borenstein, M., Hedges, L.V., Higgins, J.P. and Rothstein, H.R., 2021. *Introduction to meta-analysis*. John Wiley & Sons.

Casey, Katherine, Glennerster, Rachel and Miguel, Edward. 2012. Reshaping Institutions: Evidence on aid impacts using a preanalysis plan. *Quarterly Journal of Economics* 127(4): 1755–1812.

Ferraro, Paul J., and Pallavi Shukla. 2022. Credibility crisis in agricultural economics. *Applied Economic Perspectives and Policy* 1–17. DOI: 10.1002/aepp.13323

Ferraro, P.J. and Shukla, P., 2020. Feature—Is a replicability crisis on the horizon for environmental and resource economics? *Review of Environmental Economics and Policy* 14(2): 339-351.

Heller, R., Rosenbaum, P. R., & Small, D. S. (2009). Split samples and design sensitivity in observational studies. *Journal of the American Statistical Association*, 104 (487), 1090-1101.
http://dx.doi.org/10.1198/
jasa.2009.tm08338

Luo, X. and Deng, L., 2009. Do birds of a feather flock higher? The effects of partner similarity on innovation in strategic alliances in knowledge-intensive industries. *Journal of Management Studies*, *46*(6), pp.1005-1030.

Miguel, E., Camerer, C., Casey, K., Cohen, J., Esterling, K.M., Gerber, A., Glennerster, R., Green, D.P., Humphreys, M., Imbens, G. and Laitin, D., 2014. Promoting transparency in social science research. *Science*, *343*(6166), pp.30-31.

Sen, A. (1976). Poverty: An ordinal approach to measurement. *Econometrica*, 44(2), 219–231. https://doi.org/10.2307/1912718

Tian, Z., Wojan, T.R. and Goetz, S.J. (2022) *An examination of the informational content from self-reported innovation questions*. CES Working Paper Number CES-22-46. https://www.census.gov/library/working-papers/2022/adrm/CES-WP-22-46.html

Wojan, T.R. (2022) Ownership diversity and innovation: Using split sample design to generate more credible findings, presented at the Federal Committee on Statistical Methodology Research and Policy Conference, Washington, DC, October 25.

Appendix

Table A1 Dual Attribute Diversity and New-to-Market Innovation Failing Pass Through Threshold

Attribute	Estimate	Standard Error	Odds Ratio
MU	0.0269	0.0126	1.027
HM	-0.0244	0.0125	0.976
GU	0.314	0.0273	1.369
GR	0.1976	0.0297	1.219
GM	0.0607	0.0156	1.063
GH	0.133	0.0298	1.142
EM	0.2475	0.0147	1.281
AR	0.3522	0.0255	1.422
AM	0.075	0.0138	1.078
AH	0.3159	0.0261	1.371
AG	0.1327	0.021	1.142

Source: 2018 Annual Business Survey, 35% Exploratory Sample
Note: Coefficient estimates for firm size class and industry controls are not shown. Composite measure abbreviations include Age = A; E = Education; G = Gender; H = Ethnicity; M = Educational Specialization; R = Race; U = Foreign Born

Table A2 Three Attribute Diversity and New-to-Market Innovation Failing Pass Through Threshold

Attribute	Estimate	Standard Error	Odds Ratio
MRU	-0.0829	0.0093	0.92
HMU	-0.0877	0.0093	0.916
HMR	-0.1004	0.0093	0.904
GMU	0.086	0.0108	0.918
GMR	-0.1034	0.0107	0.902
GHM	-0.1095	0.0107	0.896
EMU	-0.0131	0.0105	0.987
EMR	-0.0306	0.0104	0.97
EHM	-0.0365	0.0104	0.964
EGM	-0.0093	0.012	0.991
AMU	-0.0667	0.0102	0.935
AMR	-0.0827	0.0102	0.921
AHM	-0.0884	0.0102	0.915
AGR	0.3128	0.0297	1.367
AGM	-0.0813	0.0118	0.922
AGH	0.2731	0.0299	1.314
AEM	0.0041	0.0112	1.004

Source: 2018 Annual Business Survey, 35% Exploratory Sample

Note: Coefficient estimates for firm size class and industry controls are not shown. Composite measure abbreviations include Age = A; E = Education; G = Gender; H = Ethnicity; M = Educational Specialization; R = Race; U = Foreign Born

Table A3 Four Attribute Diversity and New-to-Market Innovation Failing Pass Through Threshold

Attribute	Estimate	Standard Error	Odds Ratio
HMRU	-0.1043	0.0082	0.901
GMRU	-0.1111	0.0091	0.895
GHMU	-0.1144	0.0091	0.892
GHMR	-0.1229	0.009	0.884
EMRU	-0.0705	0.0089	0.932
EHMU	-0.0737	0.0089	0.929
EHMR	-0.0825	0.0089	0.921
EGMU	-0.0708	0.0099	0.932
EGMR	-0.0818	0.0098	0.921
EGHM	-0.0856	0.0098	0.918
AMRU	-0.0983	0.0088	0.906
AHMU	-0.1015	0.0088	0.903
AHMR	-0.1097	0.0087	0.896
AGMU	-0.1061	0.0098	0.899
AGMR	-0.1164	0.0097	0.89
AGHM	-0.1202	0.0097	0.887
AEMU	-0.0588	0.0095	0.943
AEMR	-0.0692	0.0095	0.933
AEHM	-0.0728	0.0095	0.93
AEGM	-0.0674	0.0105	0.935

Source: 2018 Annual Business Survey, 35% Exploratory Sample

Note: Coefficient estimates for firm size class and industry controls are not shown. Composite measure abbreviations include Age = A; E = Education; G = Gender; H = Ethnicity; M = Educational Specialization; R = Race; U = Foreign Born

Table A4 Five Attribute Diversity and New-to-Market Innovation Failing Pass Through Threshold

Attribute	Estimate	Standard Error	Odds Ratio
GHMRU	-0.1187	0.0082	0.888
EHMRU	-0.0913	0.0081	0.913
EGMRU	-0.0944	0.0088	0.91
EGHMU	-0.0968	0.0088	0.908
EGHMR	-0.1034	0.0088	0.902
AHMRU	-0.1095	0.008	0.896
AGMRU	-0.1162	0.0087	0.89
AGHMU	-0.1187	0.0087	0.888
AGHMR	-0.1249	0.0087	0.883
AEMRU	-0.0851	0.0086	0.918
AEHMU	-0.0875	0.0086	0.916
AEHMR	-0.0939	0.0085	0.91
AEGMU	-0.0889	0.0093	0.915
AEGMR	-0.0965	0.0092	0.908
AEGHM	-0.0992	0.0092	0.906

Source: 2018 Annual Business Survey, 35% Exploratory Sample
Note: Coefficient estimates for firm size class and industry controls are not shown. Composite measure abbreviations include Age = A; E = Education; G = Gender; H = Ethnicity; M = Educational Specialization; R = Race; U = Foreign Born

Table A5 Six Attribute Diversity and New-to-Market Innovation Failing Pass Through Threshold

Attribute	Estimate	Standard Error	Odds Ratio
EGHMRU	-0.1052	0.0082	0.9
AGHMRU	-0.1207	0.0081	0.886
AEHMRU	-0.0979	0.008	0.907
AEGMRU	-0.1017	0.0085	0.903
AEGHMU	-0.1036	0.0085	0.902
AEGHMR	-0.1088	0.0085	0.897

Source: 2018 Annual Business Survey, 35% Exploratory Sample
Note: Coefficient estimates for firm size class and industry controls are not shown. Composite measure abbreviations include Age = A; E = Education; G = Gender; H = Ethnicity; M = Educational Specialization; R = Race; U = Foreign Born

Table A6 Seven Attribute Diversity and New-to-Market Innovation Failing Pass Through Threshold

Attribute	Estimate	Standard Error	Odds Ratio
AEGHMRU	-0.1089	0.008	0.897

Source: 2018 Annual Business Survey, 35% Exploratory Sample

Note: Coefficient estimates for firm size class and industry controls are not shown. Composite measure abbreviations include Age = A; E = Education; G = Gender; H = Ethnicity; M = Educational Specialization; R = Race; U = Foreign Born